.tsp   XhoI   Start codon

```
GTTTTAAATACGCTCGAGGTTTTAAAT ATG TCT GTT GCC TTG TTA TGG GTT GTT TCT CCT TGT GAG
                            Met Ser Val Ala Leu Leu Trp Val Val Ser Pro Cys Asp
```

Transit peptide of phytoene synthase

```
GTC TCA AAT GGG ACA AGT TTC ATG GAA TCA GTC CGG GAG GGA AAC CGT
Val Ser Asn Gly Thr Ser Phe Met Glu Ser Val Arg Glu Gly Asn Arg
```
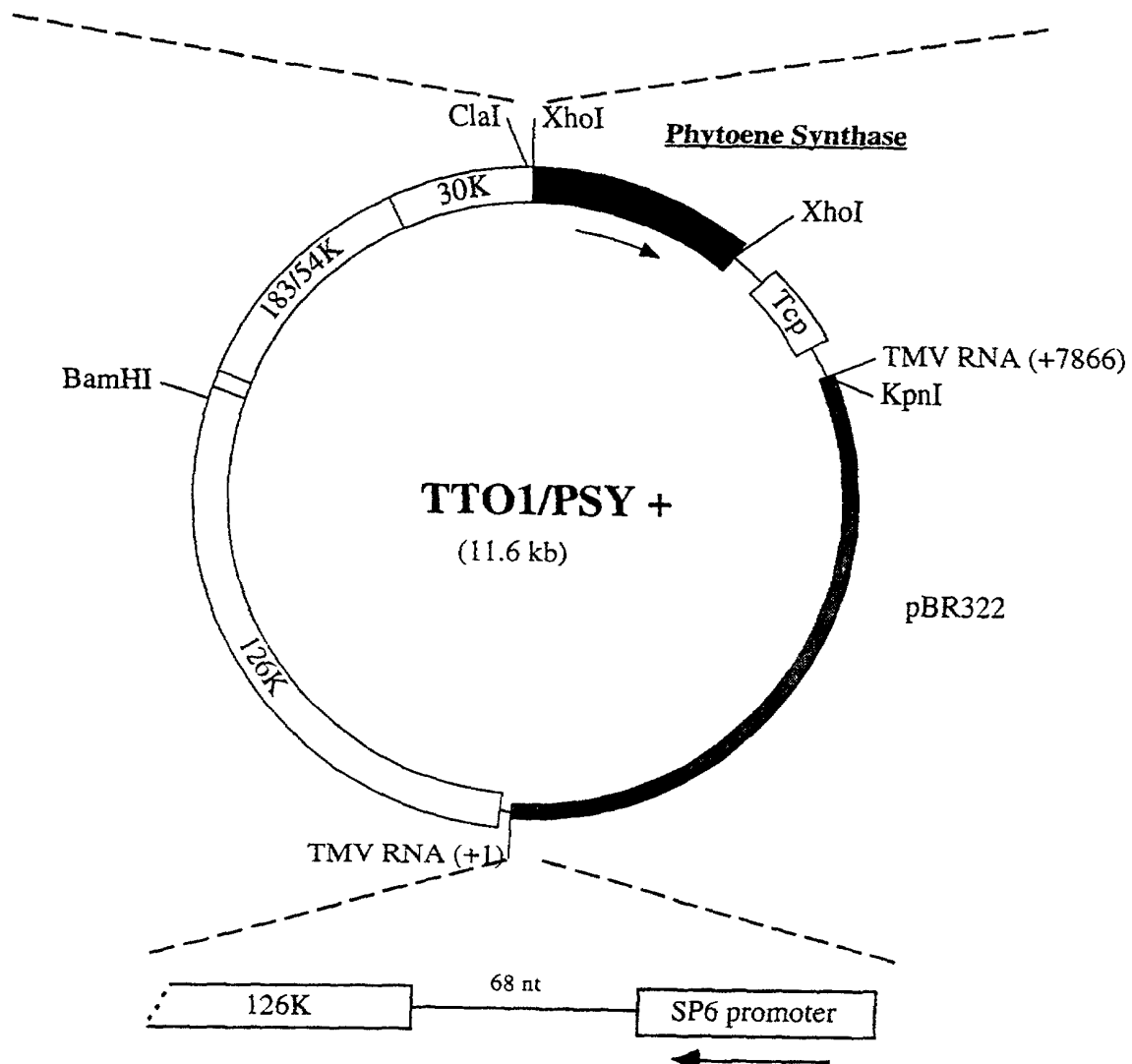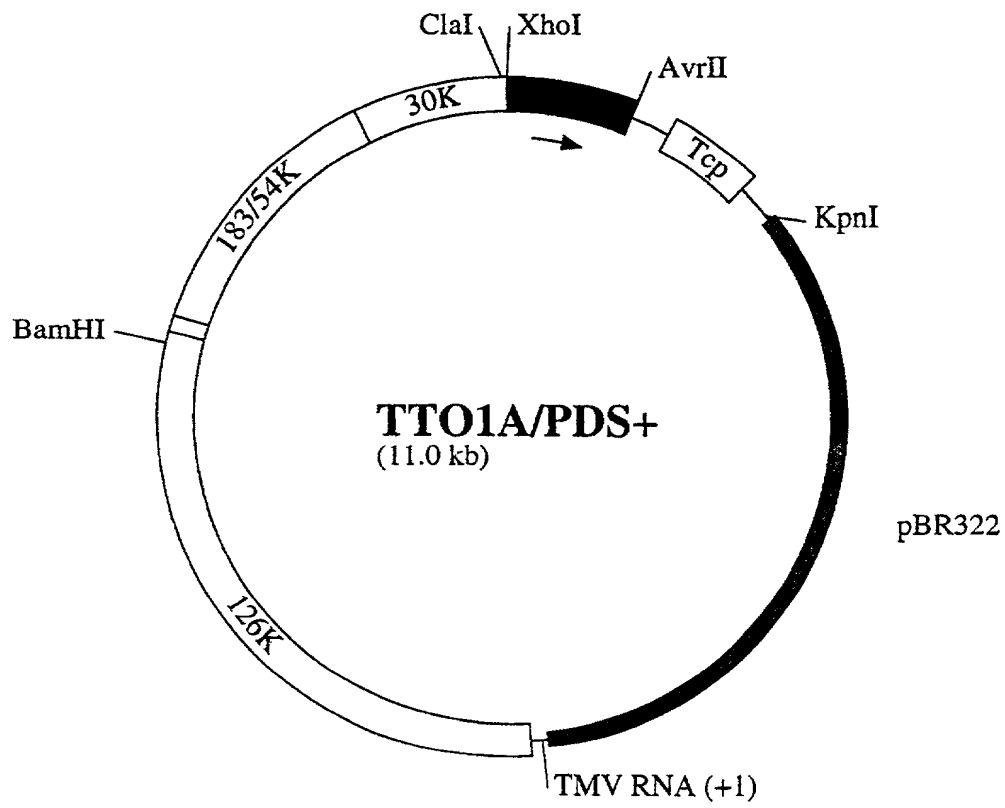

Figure 1

Figure 2

.tsp   **XhoI**   **Start codon**

```
GTTTTAAATACGCTCGAGTGCAGC ATG GAA ACC CTT CTA AAG CCT TTT CCA TCT CCT TTA CTT
                         Met Glu Thr Leu Leu Lys Pro Phe Pro Ser Pro Leu Leu
```

Transit peptide of capsanthin-capsorubin synthase

```
TCC ATT CCT ACT CCT AAC ATG TAT AGT TTC AAA CAC AAC TTC ACT TTT
Ser Ile Pro Thr Pro Asn Met Tyr Ser Phe Lys His Asn Ser Thr Phe
```
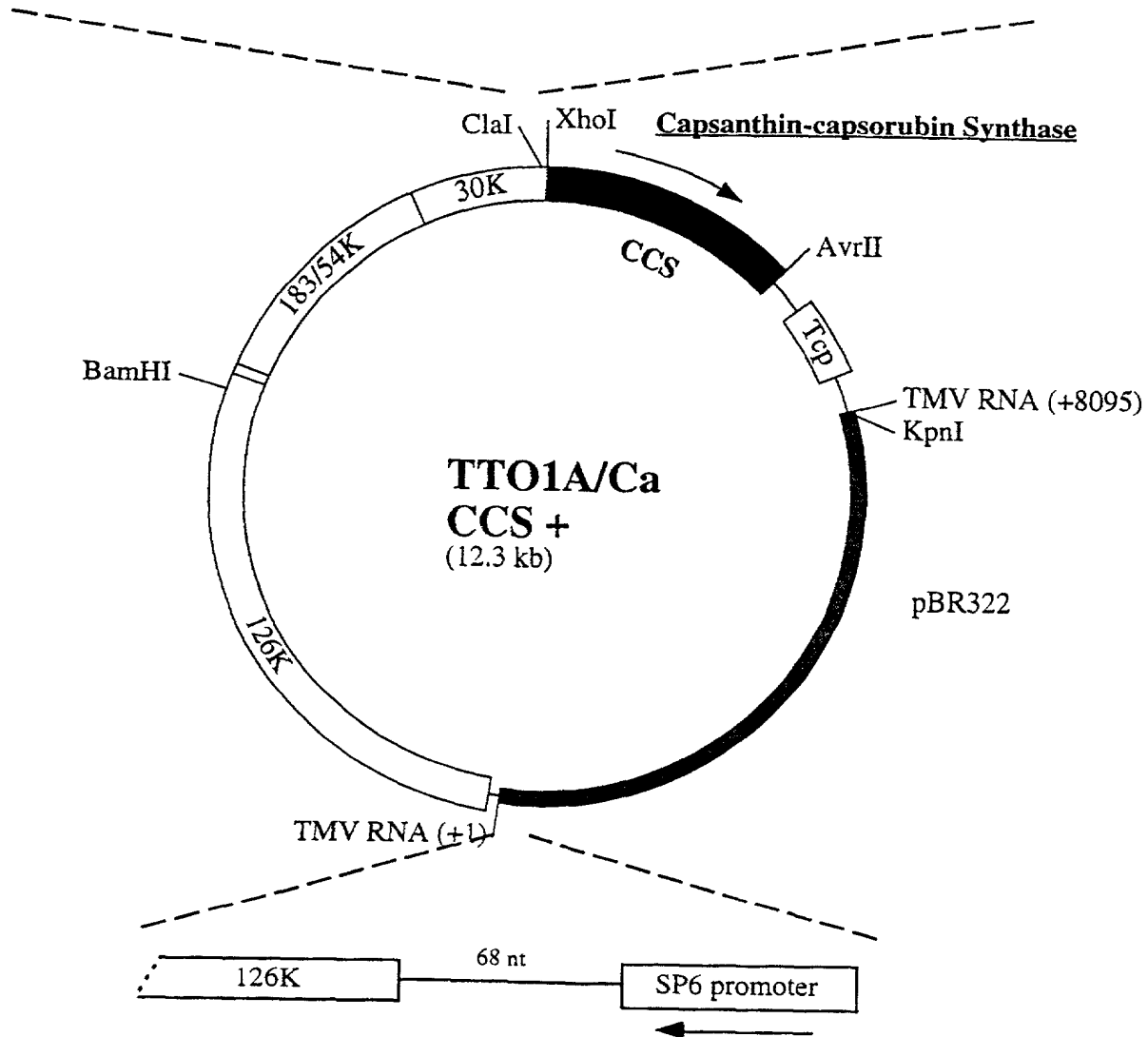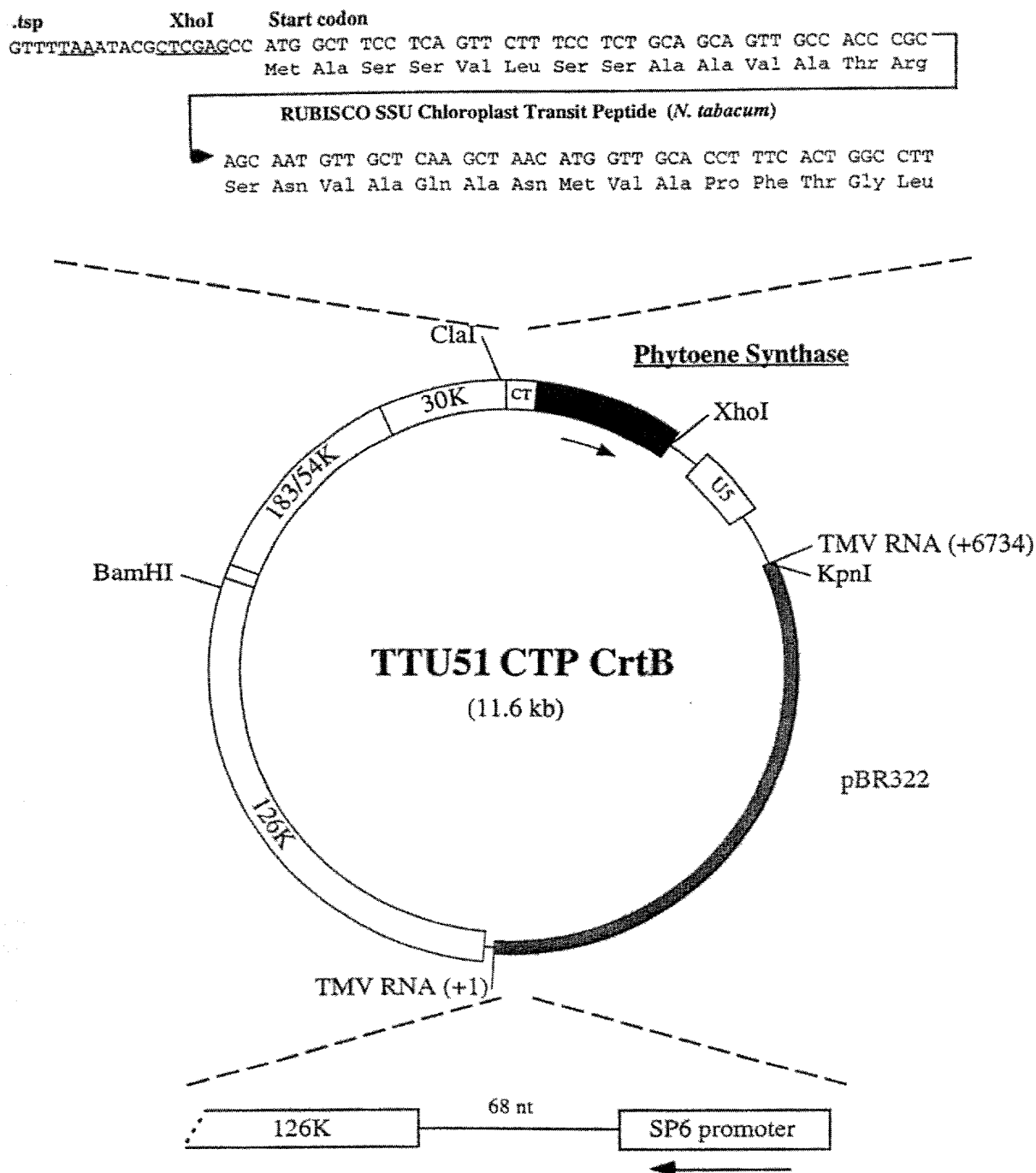

Figure 3

.tsp   XhoI Start codon
GTTT<u>TAA</u>ATACG<u>CTCGAG</u>CC ATG GCT TCC TCA GTT CTT TCC TCT GCA GCA GTT GCC ACC CGC
           Met Ala Ser Ser Val Leu Ser Ser Ala Ala Val Ala Thr Arg

RUBISCO SSU Chloroplast Transit Peptide (*N. tabacum*)

AGC AAT GTT GCT CAA GCT AAC ATG GTT GCA CCT TTC ACT GGC CTT
Ser Asn Val Ala Gln Ala Asn Met Val Ala Pro Phe Thr Gly Leu

Figure 4

Figure 5

Figure 6

225 Dra III

653 Ecl136 II
653 Sac I
658 BstX I
661 Dsa I
661 Sac II
669 Not I
670 Eag I
677 Xba I
683 Spe I
707 EcoR I
718 Avr II
718 Sty I
783 Bsm I
815 SexA I
816 Xcm I

958 BstE II

1144 Bsg I
1188 Bgl II

1335 PflM I

1551 Bsu36 I

1656 Stu I
1726 Afl II
1757 Nru I
1849 Aat II
1916 Bbs I
2031 Bcl I
2139 Sse8387 I
2199 Sph I
2208 Hind III
2214 BspD I
2214 Cla I

Xmn I 4130
Sca I 4013

Bsa I 3602
Ahd I 3530

AlwN I 3053

BspLU11 I 2642
Afl III 2642

Kpn I 2244
Acc65 I 2244
Bsp120 I 2238
Apa I 2238
EcoO109 I 2237
Xho I 2229
PaeR7 I 2229
Sal I 2223
Hinc II 2223

KS+/Crt I*491

$+$

4447 base pairs
Unique Sites

Figure 7

Figure 8

Eag I 10851
Nru I 10818
Nar I 10585
Kas I 10585
Ehe I 10585
Bbe I 10585

256 Sma I
256 Xma I
445 Sph I
811 Mlu I
981 SexA I
1130 EcoN I

1675 Stu I

2144 Ban II
2144 Ecl136 II
2144 Sac I

pBS712

11161 base pairs
Unique Sites

2650 Sac II

3177 Dra III

Ahd I 8422
Ase I 8251
Pst I 8181

3988 Xcm I

Kpn I 7424
Asp718 7424
Acc65 I 7424

BstE II 6830
Avr II 6552

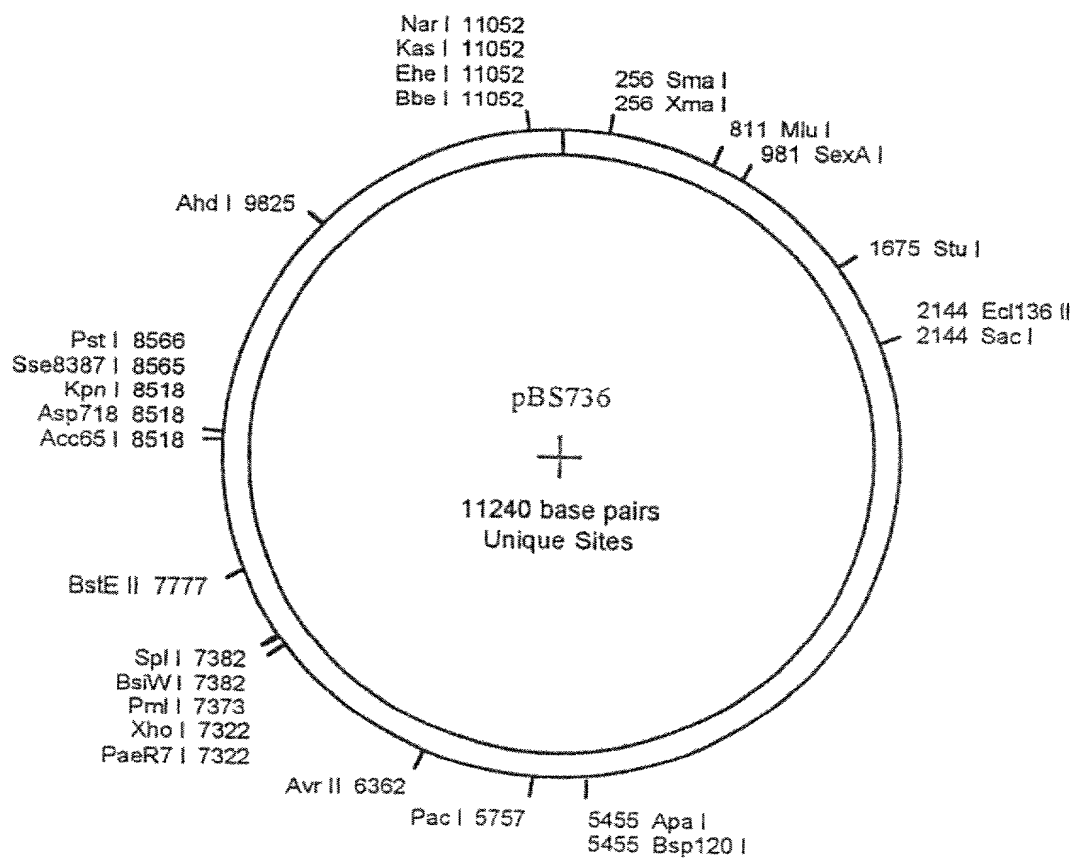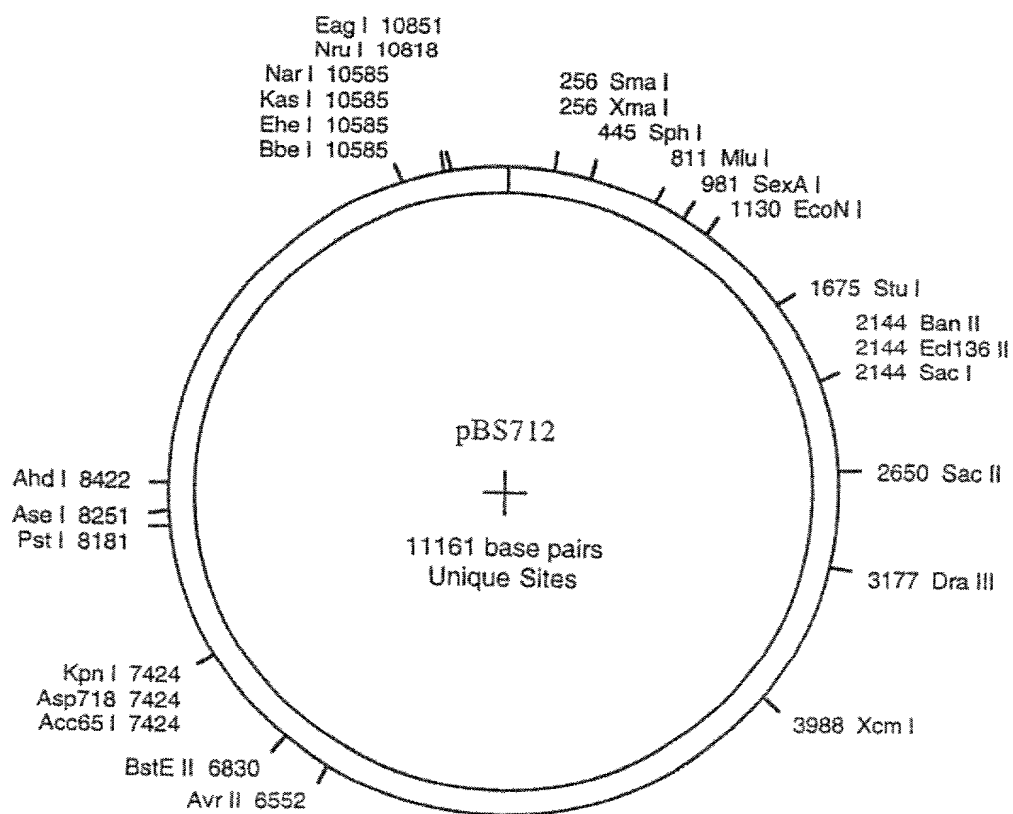**Figure 9**

```
ATGGCTATTCCCGAAGAATTTGATATTATCGTCTGTGGTGGTGGATCCAGTGGATCCTGTATTGCCGGAAGATTGGCTAACTTGGATCAC          90
METAlaIleProGluGluPheAspIleIleValCysGlyGlyGlySerSerGlySerCysIleAlaGlyArgLeuAlaAsnLeuAspHis

TCCTTGAAAGTTGGTCTTATCGAAGCAGGTGAGAACAACCTCAACAACCCATGGGTTTACCTTCCAGGTATTTACCCAAGAAACATGAAG          180
SerLeuLysValGlyLeuIleGluAlaGlyGluAsnAsnLeuAsnAsnProTrpValTyrLeuProGlyIleTyrProArgAsnMETLys

TTGGACTCCAAGACTGCCTCTTTCTACACTGCCAACCCTTCTCCTCACTTGAACGGTAGAAGGGCTATTGTCCCATGTGCTAACATCTTG          270
LeuAspSerLysThrAlaSerPheTyrThrAlaAsnProSerProHisLeuAsnGlyArgArgAlaIleValProCysAlaAsnIleLeu

GGTGGTGGTTCTTCTATCAACTTCATGATGTACACCAGAGGTTCTGCTTCTGATTATGACGACTTAGAGGCTGAAGGATGGAAAACCAAG          360
GlyGlyGlySerSerIleAsnPheMETMETTyrThrArgGlySerAlaSerAspTyrAspAspLeuGluAlaGluGlyTrpLysThrLys

GACTTGCTTCCATTGATGAAAAAGACCGAGACCTACCAAAGAGCTTGCAACAACCCTGATATCCACGGTTTTGAGGGTCCAATCAAGGTT          450
AspLeuLeuProLeuMETLysLysThrGluThrTyrGlnArgAlaCysAsnAsnProAspIleHisGlyPheGluGlyProIleLysVal

TCTTTCGGTAACTACACCTACCCAGTCTGTCAAGATTTCTTGAGAGCAAGTGAATCCCAAGGTATTCCTTACGTTGATGACTTGGAAGAC          540
SerPheGlyAsnTyrThrTyrProValCysGlnAspPheLeuArgAlaSerGluSerGlnGlyIleProTyrValAspAspLeuGluAsp

TTGGTTACCGCTCATGGTGCTGAGCACTGGTTGAAGTGGATCAACAGAGACACTGGTCGTCGTTCCGACTCCGCCCACGCCTTTGTCCAC          630
LeuValThrAlaHisGlyAlaGluHisTrpLeuLysTrpIleAsnArgAspThrGlyArgArgSerAspSerAlaHisAlaPheValHis

TCTACCATGAGAAATCACGACAACTTGTACTTGATCTGCAACACCAAGGTTGAGAAGATCATTGTTGAAGACGGAAGAGCTGTCGGTGTT          720
SerThrMETArgAsnHisAspAsnLeuTyrLeuIleCysAsnThrLysValGluLysIleIleValGluAspGlyArgAlaValGlyVal

AAAACCGTTCCAAGCAAGCCTTTGAACCCAAAGAAGCCAAGTCACAAGATTTACCGTGCTAGAAAGCAGATCGTTTTGTCTTGTGGTACC          810
LysThrValProSerLysProLeuAsnProLysLysProSerHisLysIleTyrArgAlaArgLysGlnIleValLeuSerCysGlyThr

ATCTCTTCGCCTTTGGTCTTGCAAAGATCTGGTTTCGGTGACCCAGTTAAATTGAGAGCCGCTGGTGTTAAGCCTTTGGTTAACTTGCCA          900
IleSerSerProLeuValLeuGlnArgSerGlyPheGlyAspProValLysLeuArgAlaAlaGlyValLysProLeuValAsnLeuPro

GGTGTTGGTAGAAACTTCCAGGACCACTACTGTTTCTTCACTCCTTACAGAATCAAGCCTCAATACGAGTCCTTCGATGACTTTGTCCGT          990
GlyValGlyArgAsnPheGlnAspHisTyrCysPhePheThrProTyrArgIleLysProGlnTyrGluSerPheAspAspPheValArg

GGTGACGCTGAGATTCAAAAGAGAGTCTTTGACCAATGGTACGCCAATGGTACTGGTCCTTTGGCCACCAACGGTATTGAAGCCGGTGTC          1080
GlyAspAlaGluIleGlnLysArgValPheAspGlnTrpTyrAlaAsnGlyThrGlyProLeuAlaThrAsnGlyIleGluAlaGlyVal

AAGATCAGACCAACTCCTGAGGAAATGGCTCAAATGGACGAATCCTTCCAAGAAGGTTACAGAGAATACTTTGAAGACAAGCCAGACAAG          1170
LysIleArgProThrProGluGluMETAlaGlnMETAspGluSerPheGlnGluGlyTyrArgGluTyrPheGluAspLysProAspLys

CCAGTTATGCACTACTCTATCATTGCTGGTTTCTTCGGTGACCACACCAAGATTCCTCCTGGAAAGTACATGACCATGTTCCACTTCTTG          1260
ProValMETHisTyrSerIleIleAlaGlyPhePheGlyAspHisThrLysIleProProGlyLysTyrMETThrMETPheHisPheLeu

GAGTACCCATTCTCTAGAGGTTCTATCCACATTACCTCACCAGACCCATATGCAACTCCAGACTTTGACCCAGGTTTCATGAACGATGAA          1350
GluTyrProPheSerArgGlySerIleHisIleThrSerProAspProTyrAlaThrProAspPheAspProGlyPheMETAsnAspGlu

AGAGACATGGCTCCTATGGTTTGGTCTTACAAGAAGTCTAGAGAGACTGCCAGAAGAATGGACCACTTTGCCGGTGAAGTTACTTCTCAC          1440
ArgAspMETAlaProMETValTrpSerTyrLysLysSerArgGluThrAlaArgArgMETAspHisPheAlaGlyGluValThrSerHis

CACCCTCTGTTCCCATACTCATCTGAAGCCAGAGCTTACGAGATGGATTTGGAGACTTCCAACGCTTACGGTGGACCATTGAACTTGTCC          1530
HisProLeuPheProTyrSerSerGluAlaArgAlaTyrGluMETAspLeuGluThrSerAsnAlaTyrGlyGlyProLeuAsnLeuSer

GCTGGTCTTGCACACGGTTCTTGGACTCAACCTTTGAAGAAGCCAACTGCCAAGAACGAAGGTCACGTTACCTCCAACCAGGTTGAACTT          1620
AlaGlyLeuAlaHisGlySerTrpThrGlnProLeuLysLysProThrAlaLysAsnGluGlyHisValThrSerAsnGlnValGluLeu

CACCCAGACATCGAGTACGATGAGGAGGATGACAAGGCCATTGAGAACTACATCCGTGAGCACACTGAGACCACATGGCACTGTCTGGGA          1710
HisProAspIleGluTyrAspGluGluAspAspLysAlaIleGluAsnTyrIleArgGluHisThrGluThrThrTrpHisCysLeuGly

ACCTGTTCCATCGGTCCAAGAGAAGGTTCCAAGATTGTTAAATGGGGTGGTGTTCTAGACAACAGATCCAACGTTACGGAGTCAAGGGC          1800
ThrCysSerIleGlyProArgGluGlySerLysIleValLysTrpGlyGlyValLeuAspAsnArgSerAsnValTyrGlyValLysGly

TTGAAGGTTGGTGACTTGTCTGTTTGTCCAGACAATGTTGGTTGTAACACTTACACTACCGCTCTTTTGATTGGTGAAAAGACTGCCACT          1890
LeuLysValGlyAspLeuSerValCysProAspAsnValGlyCysAsnThrTyrThrThrAlaLeuLeuIleGlyGluLysThrAlaThr

CTGGTTGGTGAAGACTTAGGATACTCTGGTGAGGCCTTAGACATGACTGTTCCACAATTCAAGTTGGGTACTTATGAGAAGACAGGTCTT          1980
LeuValGlyGluAspLeuGlyTyrSerGlyGluAlaLeuAspMETThrValProGlnPheLysLeuGlyThrTyrGluLysThrGlyLeu

GCTAGATTCTAA                                                                                    1992
AlaArgPheSTP
```
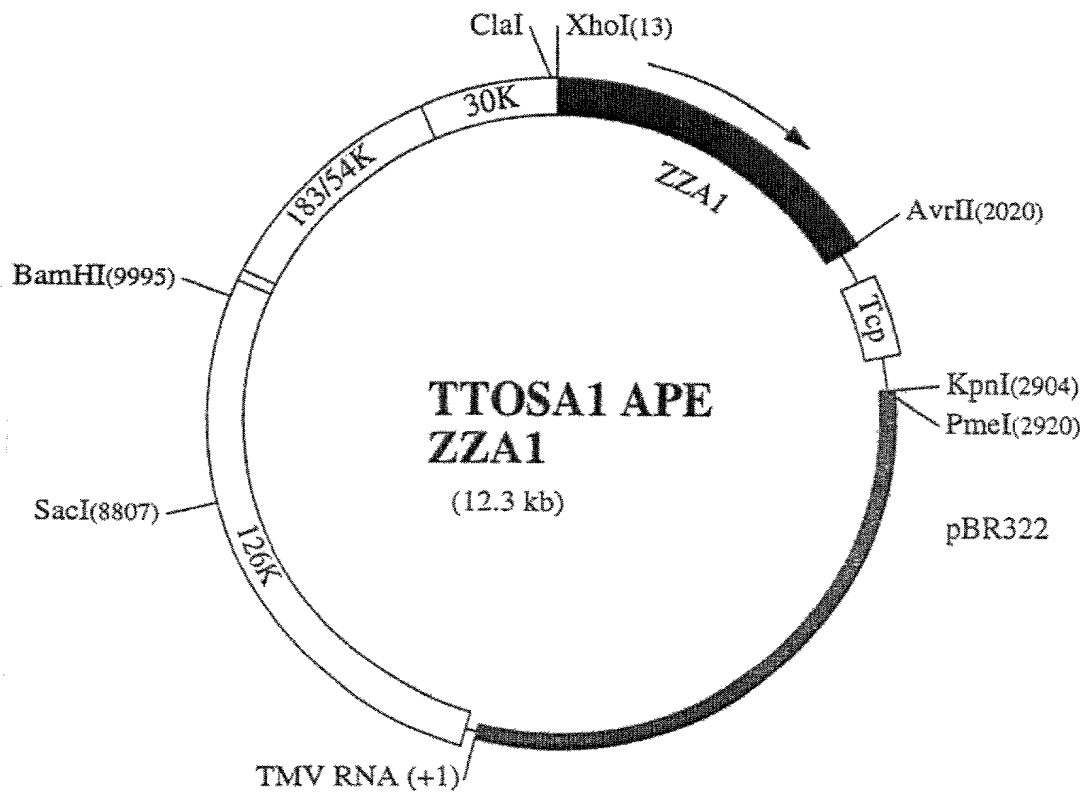
Figure 10

Figure 11

.tsp   XhoI   Start codon
GTTTTAAATACGCTCGAGATCAATCCATCTCCGAAGTGTGTCTGCAGC ATG CAG GTG CTG AAC ACC ATG
                                                 Met Gln Val Leu Asn Thr Met

                    Rice α-amylase signal peptide
    GTG AAC AAA CAC TTC TTG TCC CTT TCG GTC CTC ATC GTC CTC ATC GTC
    Val Asn Lys His Phe Leu Ser Leu Ser Val Leu Ile Val Leu Leu Gly

Figure 12

.tsp   XhoI   Start codon

GTTTTAAATACGCTCGAGG ATG ATC AGA TTC TTA GTC CTC TCT TTG CTA ATT CTC ACC CTC TTC
                           Met Ile Arg Phe Leu Val Leu Ser Leu Leu Ile Leu Thr Leu Phe
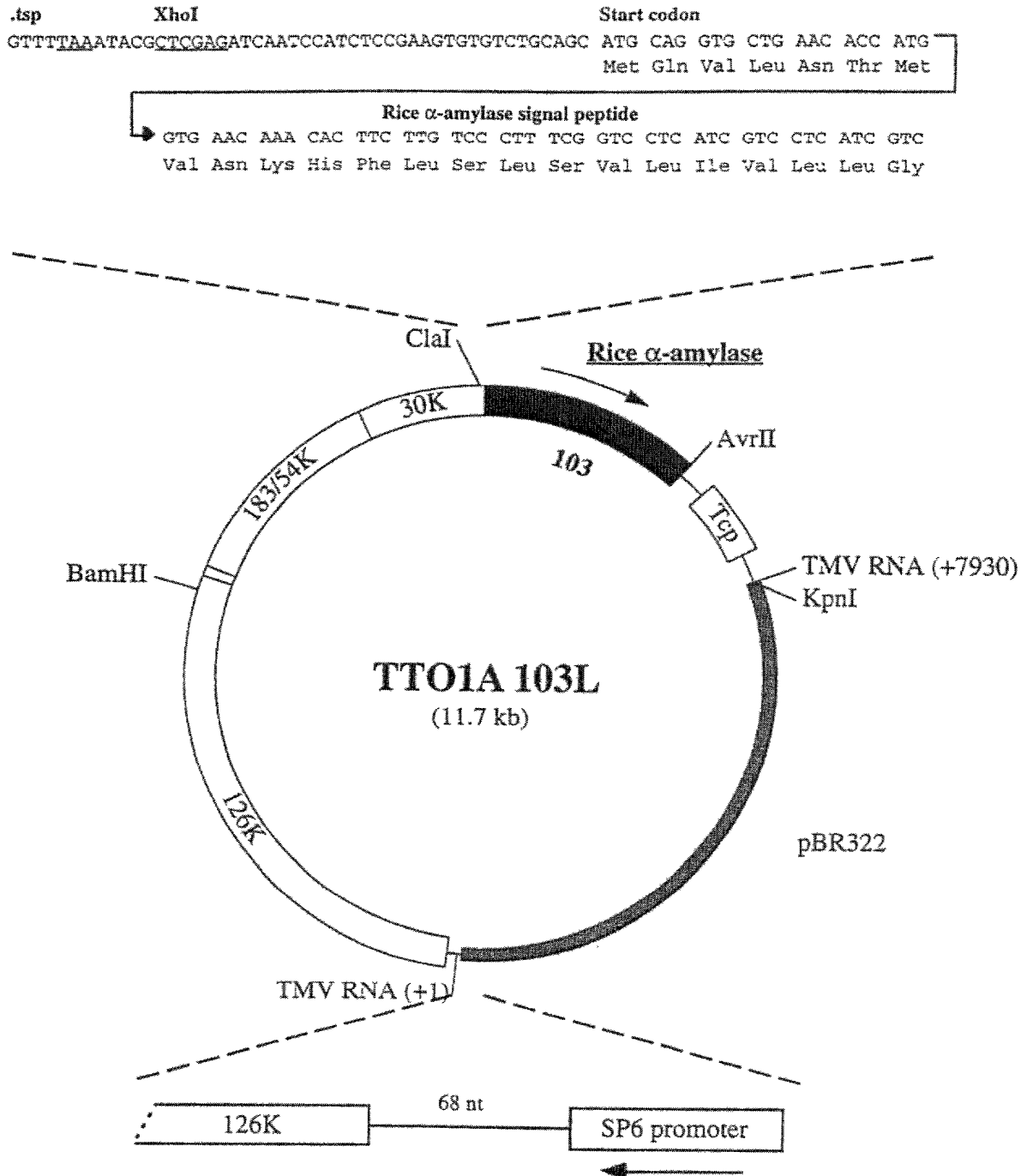
               Signal peptide   -1   +1   Mature α-trichosanthin
CTA ACA ACT CCT GCT GTG GAG GGC | GAT GTT AGC TTC CGT TTA TCA
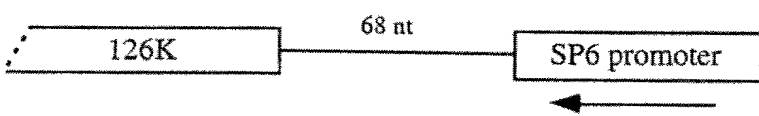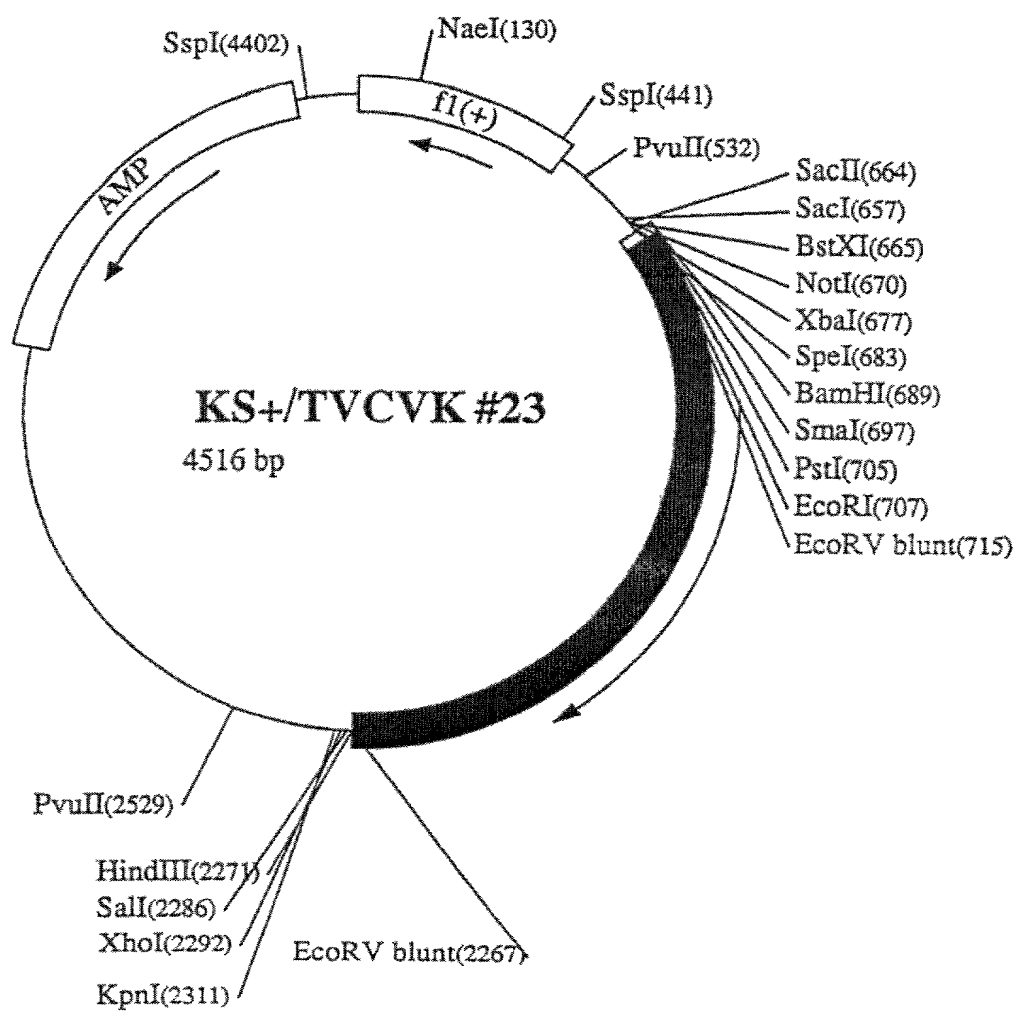Leu Thr Thr Pro Ala Val Glu Gly | Asp Val Ser Phe Arg Leu Ser

TTU51A QSEO #3
(11.5 kb)

ClaI   XhoI   α-trichosanthin
30K
AvrII
183/54K
U5cp
BamHI
KpnI
126K
pBR322
TMV RNA (+1)

126K   68 nt   SP6 promoter

Figure 13

Figure 14

Nar I 10596
Kas I 10596
Ehe I 10596
Bbe I 10596

308 Bsm I

811 Mlu I

1130 EcoN I

1675 Stu I

1922 Nhe I
2144 Ecl136 II
2144 Sac I

pBS735

10784 base pairs
Unique Sites

Sse8387 I 8109
Kpn I 8062
Asp718 8062
Acc65 I 8062

BstE II 7321

Spl I 6926
BsiWI 6926
Xho I 6866
PaeR7 I 6866
Eag I 6837
Sgf I 6665

Pac I 5757
Bsp120 I 5455
Apa I 5455

Figure 15

Figure 16

220 Dra III
223 BsaA I
326 Nae I

653 BstX I
656 Sac II
664 Not I
665 Eag I
672 Xba I
678 Spe I
691 Sma I
691 Xma I
809 BstB I
827 Hinc II
827 Sal I
870 Mun I

982 Bst1107 I
982 Sna I
982 Xca I
1108 Mlu I
1154 BsrG I
1201 Nde I
1208 BspE I
1261 Msc I
1309 Hind III

1437 Sph I
1439 Nsi I
1522 Afl II
1584 Pst I
1629 Acc65 I
1629 Asp718
1629 Kpn I

Xmn I 3515
Aha II 3456
Sca I 3398

Fsp I 3140

Bsa I 2987

Ahd I 2915

AlwN I 2438

pBS723

3838 base pairs
Unique Sites

Figure 17

Figure 18

BglII(3)

NcoI(10137)
NotI(53)
NotI(753)
EagI(754)
XhoI(760)

30K

U5 cp

3·0

KpnI(1956)
PstI(2004)

18/54K

BamHI

pBS740 AT
#120   (10.4 kb)

pUC19

SacI(6822)

126K

TMV RNA (+1)
SmaI(4934)

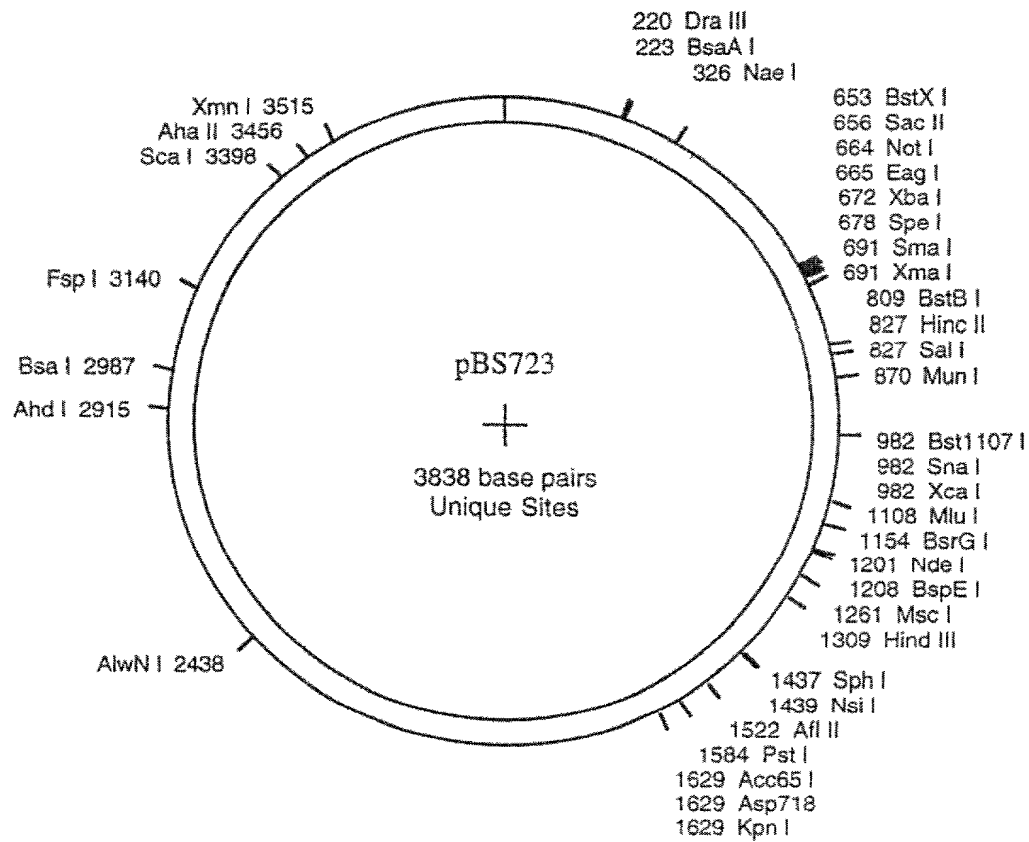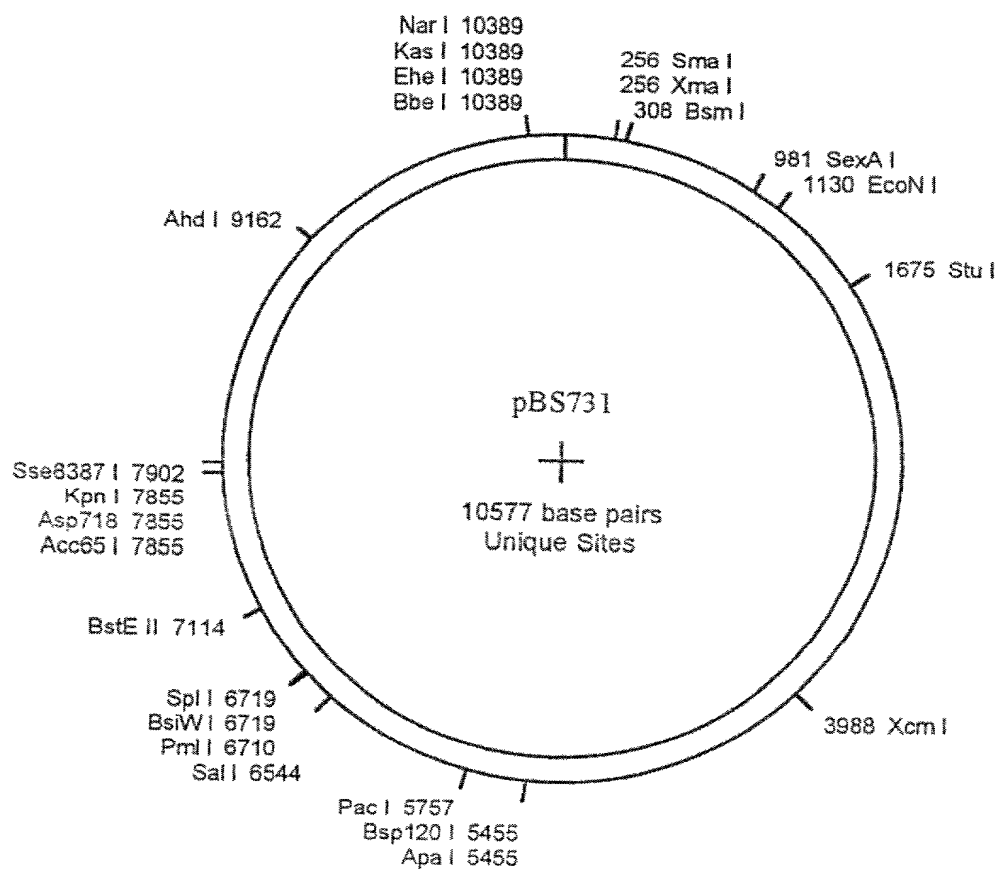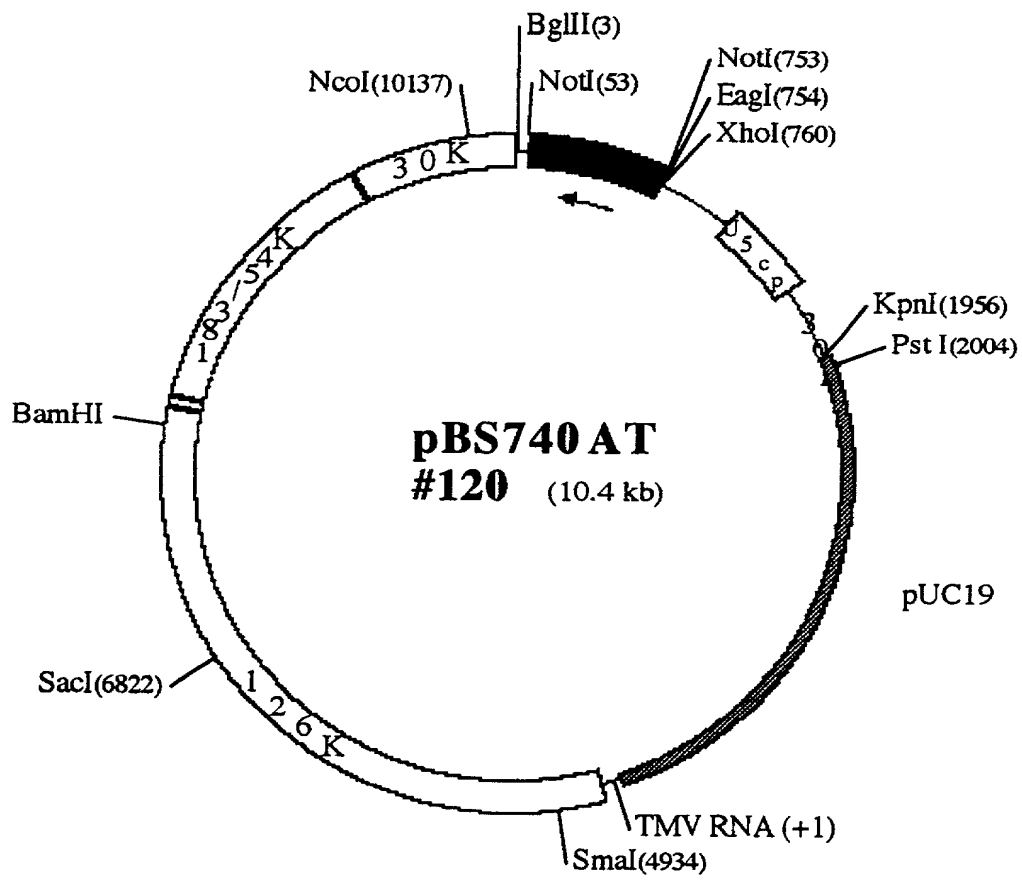**Figure 19**

Nucleotide sequence alignment of 740 AT #120 to human ADP-ribosylation factor
(ARF3) M33384

```
740 AT #120   AAGAAGGAGATGCGAATTCTGATGGTTGGTCTTGATGCTGCTGGTAAGACCACAATCTTG
              ||||||||||||| || ||||||||| || || ||||| || || |||||||| ||| |
M33384        AAGAAGGAGATGCGCATCCTGATGGTGGGCCTGGATGCCGCAGGAAAGACCACCATCCTA
              K   K   E   M   R   I   L   M   V   G   L   D   A   A   G   K   T   T   I   L

740 AT #120   TACAAGCTCAAGCTCGGAGAGATTGTCACCACCATCCCTACTATTGGTTTCAATGTGGAA
              ||||||||| || || || ||||| |||||||||||||||||||| ||||| |||||||||||
M33384        TACAAGCTGAAACTGGGGGAGATCGTCACCACCATCCCTACCATTGGGTTCAATGTGGAG
              Y   K   L   K   L   G   E   I   V   T   T   I   P   T   I   G   F   N   V   E

740 AT #120   ACTGTGGAATACAAGAACATTAGTTTCACCGTGTGGGATGTCGGGGGTCAGGACAAGATC
              || |||||| || ||||||||| || || || ||||||||||| || || |||||||||||
M33384        ACAGTGGAGTATAAGAACATCAGCTTTACAGTGTGGGATGTGGGTGGCCAGGACAAGATT
              T   V   E   Y   K   N   I   S   F   T   V   W   D   V   G   G   Q   D   K   I

740 AT #120   CGTCCCTTGTG-AGACACTACTTCCAGAACACTCAAGGTCTAATCTTTGTTGTTGATAGC
              || ||| | || |||||||||||||||||||||||| ||||| | || ||||| || || |||
M33384        CGACCCCTCTGGAGACACTACTTCCAGAACACCCAAGGGTTGATATTTGTGGTCGACAGC
              R   P   L   W   R   H   Y   F   Q   N   T   Q   G   L   I   F   V   V   D   S

740 AT #120   AATGACAGAGACAGAGTTGTTGAGGCTCGAGATGAACTCCACAGGATGCTGAATGAGGAC
              ||||| | || |||| |||||| || || || || || |||||| ||||||
M33384        AATGATCGGGAGCGAGTAAATGAGGCCCGGGAAGAGCTGATGAGAATGCTGGCGGAGGAC
              N   D   R   E   R   V   N   E   A   R   E   E   L   M   R   M   L   A   E   D

740 AT #120   GAGCTGCGTGATGCTGTGTTGCTTGTGTTTGCCAACAAGCAAGATCTTCCAAATGCTATG
              ||||| || ||||||||| | ||||| ||||| ||||| || ||||| || ||||||||||
M33384        GAGCTCCGGGATGCTGTACTCCTTGTCTTTGCAAACAAACAGGATCTGCCTAATGCTATG
              E   L   R   D   A   V   L   L   V   F   A   N   K   Q   D   L   P   N   A   M

740 AT #120   AACGCTGCTGAAATCACAGATAAGCTTGGCCTTCACTCCCTCCGTCAGCGTCATTGGTAT
              ||||||||||| ||||||||| ||||| ||||| || ||||| ||||| ||| | |||||
M33384        AACGCTGCTGAGATCACAGACAAGCTGGGCCTGCATTCCCTTCGTCACCGTAACTGGTAC
              N   A   A   E   I   T   D   K   L   G   L   H   S   L   R   H   R   N   W   Y

740 AT #120   ATCCAGAGCACATGTGCCACTTCAGGTGAAGGGCTTTATGAAGGTCTGGACTGGCT
              || ||| ||| ||||||||| || || ||||| || ||||| |||||||||||||||
M33384        ATTCAGGCCACCTGTGCCACCAGCGGGGACGGGCTGTACGAAGGCCTGGACTGGCT
              I   Q   A   T   C   A   T   S   G   D   G   L   Y   E   G   L   D   W   L
```
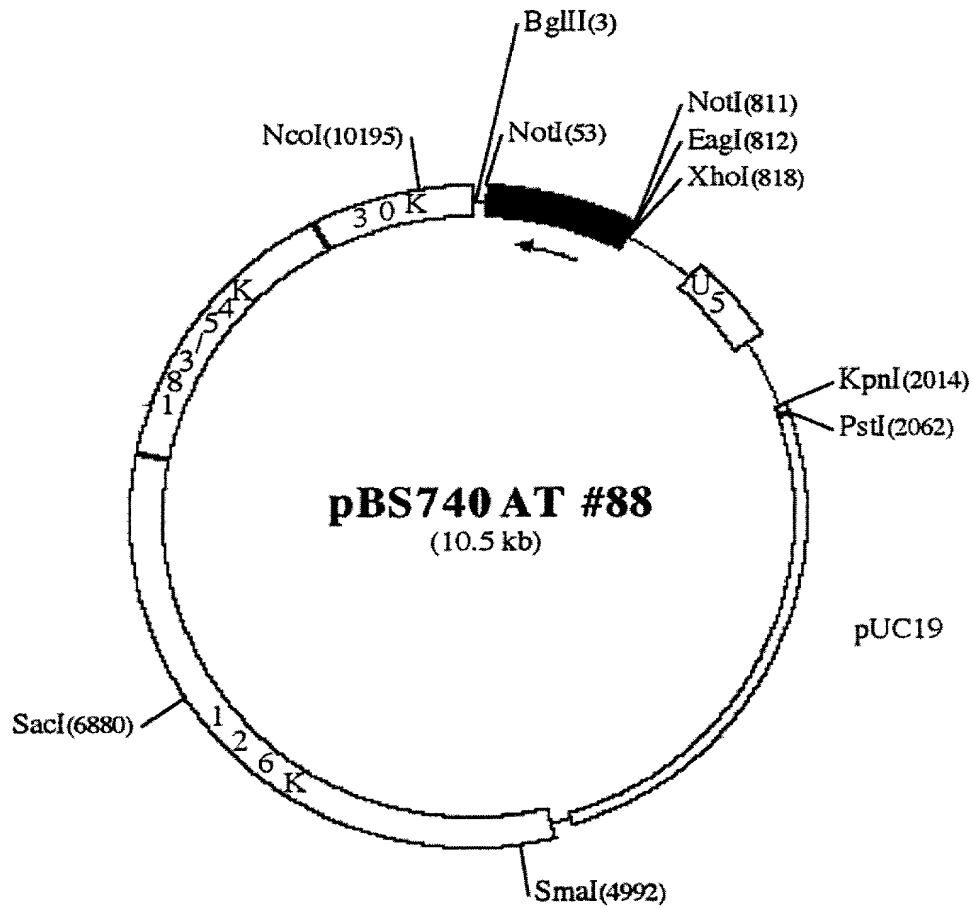
Figure 20

Figure 21

Nucleotide sequence alignment of 740 AT #88 to L33574 mRNA for rhodopsin

```
740 AT #88    CAACCTCCCGTTGGTGTTCCTCCTCCTCAAGGTTATCCACCGGAGGGATATCCAAAAGAT
              |||||||||||| ||||||||||||||||||||||||| ||||||| ||| |||||||||
L33574        CAACCTCCCGTCGGTGTTCCTCCTCCTCAAGGTTATCCCCCGGAGGATTATTCAAAAGAT

740 AT #88    GCTTATCCACCACAAGGATATCCTCCTCAGGGATATCCTCAGCAAG GCTATCCACCTCA
              |||||||||||| |||||||||||||||||||||||||||| ||||||| ||| |||||||||
L33574        GCTTATCCACCGCAAGGATATCCTCCTCAGGGCTATCCGCCGCAAGTACCCTCCACAGCA

740 AT #88    GGGATATCCTCAACAAGGTTATCCTCAGCAAGGATATCC
              |||||||||| | ||||| || |||| ||||| |||||
L33574        GGGATATCCGCCGCAAGG TACCCTCCACAAGGCTATCC
```

Identities = 45/57 (78%), Positives = 45/57 (78%)

Figure 22

Nucleotide sequence alignment of 740 AT #88 to X07797 Octopus mRNA for rhodopsin

```
740 AT #88    CCACCACAAGGATATCCTCCTCAGGGATATCCTCAGCAAGGCTATCCACCTCAGGGA
              ||||||||||||  ||  ||  ||  ||  ||  ||  ||  |  |||||||||  |||||  ||||||
X07797        CCACCACAAGGCTACCCACCACAAGGCTACCCACCTCAAGGCTACCCACCCCAGGGA
```

Identities = 45/57 (78%), Positives = 45/57 (78%)

Figure 23

G-protein-coupled receptor

N. benthamiana

```
740 AT #88    YPPQ-GYPPQGYPQQGYPPQGYPQQGYPQQGYPPPYAPQYPPPPQASATTEQVLA
              |||  |||| |||  |||| ||  |||| ||||||      |      |    ||
ATTS2938      YPPKDGYPPAGYPPAGYPPPGY-AQGYPEQGYPPPQYSQAPEEKQNAGMLEGCLA
```

Figure 24

G-protein-coupled receptor

N. benthamiana

```
740 AT #88        PPVGV-PPPQGYPPEGYPKDAYPPQGYPPQGYPQQGYPPQGYPQQGYPQQGY
                  ||  |   ||||||||||+|||    |||||||||| |  |    ||| |  ||   | ||
octopus rhodopsin PPQGAYPPPQGYPPQGYPPQGYPPQGYPPQGAPPQVEAPQGAPPQGVDNQAY
```

Identity = 34/54 65%, Posititives 35/52 66%

Figure 25

Amino acid sequence comparison of 740 AT #2441 to tobacco RAN-B1 GTP binding protein

```
Nt RAN-B1      MALPNQQTVDYPSFKLVIVGDGGTGKTTFVKRHLTGEFEKKYEPTIGVEVHPLDFFTNCG
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
740 AT #2441   MALPNQQTVDYPSFKLVIVGDGGTGKTTFVKRHLTGEFEKKYEPTIGVEVHPLDFFTNCG


Nt RAN-B1      KIRFYCWDTAGQEKFGGLRDGYYIHGQCAIIMFDVTSTTDIQECSNMAP*SLQGL*KHSQ
               |||||||||||||||||||||||||||||||||||||||+     +      |  + ++
740 AT #2441   KIRFYCWDTAGQEKFGGLRDGYYIHGQCAIIMFDVTARLTYKNVPTWHR-DLCRVCENIP


Nt RAN-B1      LFFVGIKLM*KNRQVKAQ
               +    | |+  |||||||+
740 AT #2441   IVLCGNKVDVKNRQVKAK
```
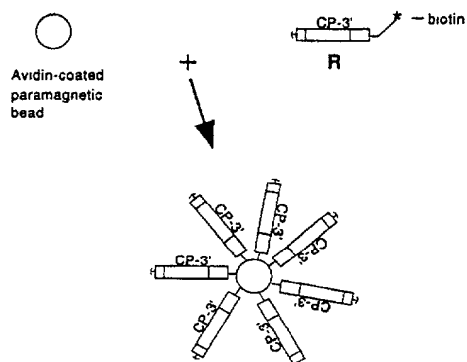
Figure 26

Nucleotide sequence comparison of 740 AT #2441 to human RAN GTP-binding protein
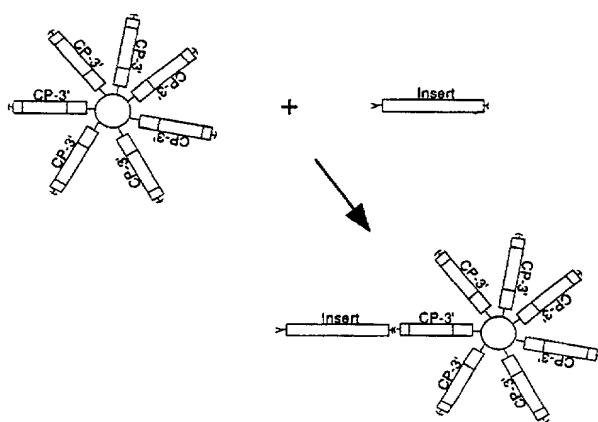
```
Human RAN     ACTGGAGAGTTTGAGAAGAAGTATGAACCCACTATTGGTGTTGAGGTTCATCC
              |||||  ||  ||||||||||||||||  |  ||||   |  ||||||||||||||||
740 AT #2241  ACTGGTGAATTTGAGAAGAAGTATGTAGCCACCTTGGGTGTTGAGGTTCATCC
```

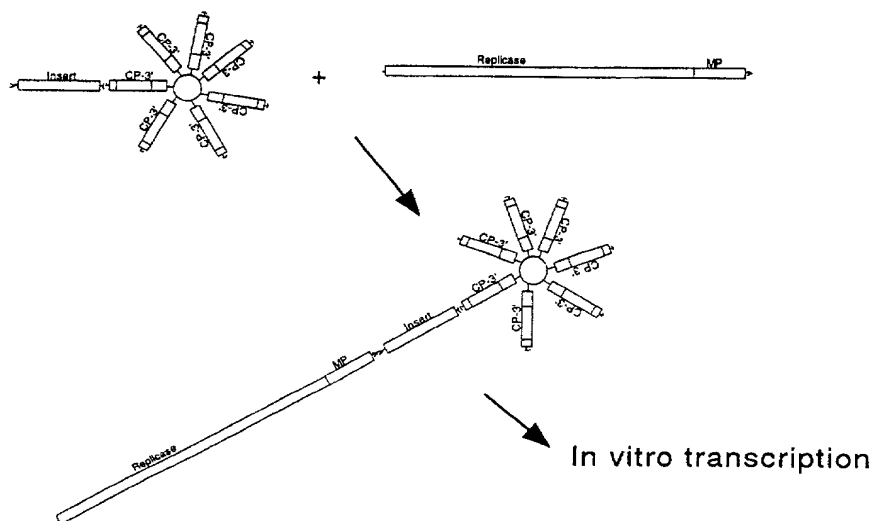Identities = 46/53 (86%), Positives = 46/53 (86%)

Figure 27

Figure 28